

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-69420

(P2000-69420A)

(43) 公開日 平成12年3月3日 (2000.3.3)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テーマコード* (参考)
H 0 4 N	5/91	H 0 4 N	N 5 C 0 2 3
	5/268		5 C 0 5 2
	5/76		B 5 C 0 5 3

審査請求 未請求 請求項の数 9 O L (全 16 頁)

(21) 出願番号 特願平10-239859

(22) 出願日 平成10年8月26日 (1998.8.26)

(71) 出願人 000005049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(72) 発明者 佐藤 克彦

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

(72) 発明者 赤木 宏之

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

(74) 代理人 100103296

弁理士 小池 隆彌

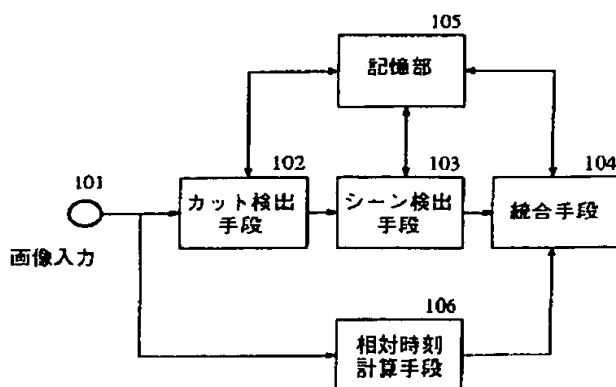
最終頁に続く

(54) 【発明の名称】 映像処理装置

(57) 【要約】

【課題】 長い映像を意味のある単位で短く統合できなかった。

【解決手段】 映像のショット間の切れ目を検出するカット検出手段102と、ショット間の画像類似度によりショットを統合するシーン検出手段103と、映像の先頭フレームからの相対時刻を計算する相対時刻計算手段106と、シーン検出手段103により検出された統合されたショット群を相対時刻計算手段106により検出された相対時刻に基づいて統合する統合手段104を有する。



**【特許請求の範囲】**

**【請求項1】** 順次入力される映像のフレームを統合する映像処理装置であって、  
フレーム間の特徴量変化に基づいてショットのカットを検出するカット検出手段と、  
前記検出されたカットにより区別された各ショットのフレーム間の類似度に基づいてシーンチェンジを検出するシーン検出手段と、  
統合のための基準フレームから入力フレームまでの相対時刻を求める相対時刻計算手段と、  
前記相対時刻が指定された一定時間となるたびに、前記検出されたシーンチェンジを区切りとしてシーン統合する統合手段と、を有することを特徴とする映像処理装置。

**【請求項2】** 順次入力される映像のフレームを統合する映像処理装置であって、  
フレーム間の特徴量変化に基づいてショットのカットを検出するカット検出手段と、  
統合のための基準フレームから入力フレームの相対時刻を求める相対時刻計算手段と、  
前記検出されたカットの基準フレームからの前記相対時刻が指定された一定時間の整数倍である場合に、CMを検出するCM検出手段と、  
前記検出されたCMを区切りとしてショット統合する統合手段と、を有することを特徴とする映像処理装置。

**【請求項3】** 映像に同期して入力される音声に基づいて順次入力される映像のフレームを統合する映像処理装置であって、  
フレーム間の特徴量変化に基づいてショットのカットを検出するカット検出手段と、  
前記検出されたカットにより区別されたショット間の入力音声の類似度に基づいて検出されたシーンチェンジを区切りとしてシーン統合する統合手段と、を有することを特徴とする映像処理装置。

**【請求項4】** 順次入力される映像に関する特徴を抽出する特徴検出手段と、  
前記抽出された特徴を前記カット検出手段により検出されたカットにより区別されたショットごとに記憶する記憶手段と、  
抽出された特徴に基づいて前記記憶手段のショットを指定されたショット数以下まで分割する分割手段と、を有することを特徴とする請求項1乃至3記載の映像処理装置。

**【請求項5】** 順次入力される映像に関する特徴を抽出する特徴検出手段と、  
前記抽出された特徴を前記カット検出手段により検出されたカットにより区別されたショットごとに記憶する記憶手段と、  
検索条件として設定された特徴に基づいて前記記憶手段からショットを検索する検索手段と、を有することを特

徴とする請求項1乃至3記載の映像処理装置。

**【請求項6】** 前記特徴検出手段が抽出する特徴は、フレーム間の動きベクトルから得られる映像中のカメラの動き情報であることを特徴とする請求項4乃至5記載の映像処理装置。

**【請求項7】** 前記特徴検出手段が抽出する特徴は、音量情報であることを特徴とする請求項4乃至5記載の映像処理装置。

**【請求項8】** 前記特徴検出手段が抽出する特徴は、ある特定の環境下における画像モデルと比較することにより求まる類似度であることを特徴とする請求項4乃至5記載の映像処理装置。

**【請求項9】** 前記特徴検出手段が抽出する特徴は、ある特定の環境下における音声モデルと比較することにより求まる類似度であることを特徴とする請求項4乃至5記載の映像処理装置。

**【発明の詳細な説明】****【0001】**

**【発明の属する技術分野】** 本発明は、映像を自動的にまとめ、まとめられた映像の編集や検索を容易にする映像処理装置に関する。

**【0002】**

**【従来の技術】** 近年、衛星放送やケーブルテレビなどによるテレビの多チャンネル化が、進んでいる。また、家庭におけるビデオムービーの普及により、映像を録画する機会が益々増えている。そのため膨大な数の映像について内容をより早く把握し、見たい部分のみを取り出すことが必要となっている。従来は、映像を早送りしながら見たい場面を探す、この場合最初から順番に見ていく必要があり、たとえ10倍の速さで見ても、2時間の映像全体を見るのには12分もかかることになり、検索に非常に時間がかかる。

**【0003】** この問題に対し、映像を分割し、その代表画像で表示することは、映像全体を理解し、編集や検索を行なう上で非常に便利である。この分割方法として、カットによる分割が用いられる場合が多い。カットとは、カメラの切替えによって映像が大きく変化する点であり、カットと次のカットの間は連続した内容を持つ一連の画像フレームで構成されている。この一連のフレーム群をショットと呼び、映像の1つの単位として扱うことができる。しかしながら、映像をカットごとに分ける作業は、非常に大きな労力が必要であるため、自動化することが大きな課題であり、多くの技術が開発されている。

**【0004】** また、映像を意味のある内容ごとにまとめるために、ショットを統合する方法として、特開平9-93588号公報に示されているように、動画像のフレーム群間の類似度を求め、リンクを付け同一の集合として統合する方法がある。

**【0005】**

【発明が解決しようとする課題】しかしながら、映像をカット検出のみで分割する方法では、多くの映像はカットが平均約5秒～10秒に1回起こると言われているため、ショット数が多く検出され、長い映像が短くまとまらない。例えば、カットが10秒に一度起こると仮定して、30分の映像についてカット検出すると180箇所検出され、それを1画面に30枚の画像が表示できる表示装置に表示すると、画面6枚分となる。これでは、映像全体を把握することは難しい。

【0006】さらに、CM（コマーシャル）などでは、極端に短いショットが数多くあり、重要な内容を含まないショットが数多く表示されることになり、内容全体を把握することは困難である。

【0007】また、画像の類似度のみでショットを統合する方法では、統合されるショット数は多くても数個から数十個程度である。統合してもやはりショット数が多く、長い映像を短くまとめることは困難である。

【0008】また、映像中のショットには重要な内容を持つ部分とそうでない部分があり、映像を短くまとめる際、重要な内容を持つショットを残すことが必要である。

【0009】また、映像中から特定のショットや同じ背景のショットだけをまとめて検索して取り出したい、あるいは編集したいなどの要求がある。そのため、映像の様々な特徴を基に望みのショットを表示したり、実際に検索できることが必要となる。

【0010】本発明の目的は、上記問題点に鑑み、映像全体を容易に把握したり、短くまとめたり等できる映像処理装置を提供することにある。

【0011】

【課題を解決するための手段】請求項1記載の映像処理装置は、順次入力される映像のフレームを統合する映像処理装置であって、フレーム間の特徴量変化に基づいてショットのカットを検出するカット検出手段と、前記検出されたカットにより区別された各ショットのフレーム間の類似度に基づいてシーンチェンジを検出するシーン検出手段と、統合のための基準フレームから入力フレームまでの相対時刻を求める相対時刻計算手段と、前記相対時刻が指定された一定時間となるたびに、前記検出されたシーンチェンジを区切りとしてシーン統合する統合手段と、を有することを特徴とする。

【0012】請求項2記載の映像処理装置は、順次入力される映像のフレームを統合する映像処理装置であって、フレーム間の特徴量変化に基づいてショットのカットを検出するカット検出手段と、統合のための基準フレームから入力フレームの相対時刻を求める相対時刻計算手段と、前記検出されたカットの基準フレームからの前記相対時刻が指定された一定時間の整数倍である場合に、CMを検出するCM検出手段と、前記検出されたCMを区切りとしてショット統合する統合手段と、を有す

ることを特徴とする。

【0013】請求項3記載の映像処理装置は、映像に同期して入力される音声に基づいて順次入力される映像のフレームを統合する映像処理装置であって、フレーム間の特徴量変化に基づいてショットのカットを検出するカット検出手段と、前記検出されたカットにより区別されたショット間の入力音声の類似度に基づいて検出されたシーンチェンジを区切りとしてシーン統合する統合手段と、を有することを特徴とする。

【0014】請求項4記載の映像処理装置は、請求項1乃至3記載の映像処理装置において、順次入力される映像に関する特徴を抽出する特徴検出手段と、前記抽出された特徴を前記カット検出手段により検出されたカットにより区別されたショットごとに記憶する記憶手段と、抽出された特徴に基づいて前記記憶手段のショットを指定されたショット数以下まで分割する分割手段と、を有することを特徴とする。

【0015】請求項5記載の映像処理装置は、請求項1乃至3記載の映像処理装置において、順次入力される映像に関する特徴を抽出する特徴検出手段と、前記抽出された特徴を前記カット検出手段により検出されたカットにより区別されたショットごとに記憶する記憶手段と、検索条件として設定された特徴に基づいて前記記憶手段からショットを検索する検索手段と、を有することを特徴とする。

【0016】請求項6記載の映像処理装置は、請求項4乃至5記載の映像処理装置において、前記特徴検出手段が抽出する特徴は、フレーム間の動きベクトルから得られる映像中のカメラの動き情報であることを特徴とする。

【0017】請求項7記載の映像処理装置は、請求項4乃至5記載の映像処理装置において、前記特徴検出手段が抽出する特徴は、音量情報であることを特徴とする。

【0018】請求項8記載の映像処理装置は、請求項4乃至5記載の映像処理装置において、前記特徴検出手段が抽出する特徴は、ある特定の環境下における画像モデルと比較することにより求まる類似度であることを特徴とする。

【0019】請求項9記載の映像処理装置は、請求項4乃至5記載の映像処理装置において、前記特徴検出手段が抽出する特徴は、ある特定の環境下における音声モデルと比較することにより求まる類似度であることを特徴とする。

【0020】

【発明の実施の形態】まず、図2を用いて、映像の構成について説明する。本実施の形態で入力される映像は、入力時に何らかの時系列的あるいは番号的に順序付けられた画像、例えば動画画像であればよく、例えばテレビ、VTR（デジタルテープレコーダ）、ポータブルビデオカメラなどの外部から入力される信号でも良いし、コン

コンピュータのハードディスク、光磁気ディスク、CD-R OM（コンパクトディスクリードオンリーメモリ）などの記録媒体に保持されている符合化された信号でも良い。ビデオカメラで撮影した場合を例にとると、映像を構成する1枚1枚の画像を「フレーム」、撮影を中断せずに撮影された一続きの映像のフレーム群を「ショット」、その境界を「カット」と呼ぶ。また、内容が連続しているショット群を「シーン」とし、「シーン」の境界を「シーンチェンジ」と呼ぶ。さらに、連続シーンでの同一性を基に「シーン」を統合し、新しい「シーン」を構成する。この「シーン」は、階層が何段あってもよい。

【0021】（第1の実施の形態）以下に、本実施の形態を示す図面に基づき説明する。図1に、第1の実施の形態の映像処理装置の構成を示す。101は画像入力端であり、102は入力された画像を用いてカットを検出するカット検出手段であり、103はシーンチェンジを検出するシーン検出手段であり、104は106からの相対時刻でシーンを統合する統合手段であり、105はワークエリアのみならず本装置の各部で処理された映像を記憶する記憶部であり、106は先頭フレームからの相対時刻を求める相対時刻計算手段である。相対時刻検出手段106は、先頭時刻のフレーム番号と入力フレーム番号との差から計算できる。例えば、30フレーム/秒のビデオ信号であれば、フレーム番号の差が120であれば、4秒であることが分かる。

【0022】次に、カット検出手段102の処理について説明する。カット検出手段102は、入力された画像の連続するフレーム群のうち、特徴量変化の大きいフレーム間をカットとして検出するものである。これによって、連続的に変化の小さいフレーム群をショットとしてまとめる。前記特徴量変化の検出としては、たとえば、画素変化面積を用いる方法やフレーム間での輝度ヒストグラムを用いて比較する方法や、色の分布の変化を求める方法などがあるが、特に限定されない。

【0023】この特徴量変化の検出として、画素変化面積を用いる方法について図3を用いて説明する。1枚の画像は画素と呼ばれる小さな形（矩形の場合が多い）の集まりで表される。図3（a）は時間 $t$ の画像の一部分を表しているとする。時間が $t+1$ の時の同一の場所における画像が図3（b）であれば、時間 $t$ と $t+1$ の変化を表す差分画像は図3（d）となる。同様に時間 $t+1$ の画像が図3（c）であれば、差分画像は図3（e）となる。カットが起こった場合、画像が不連続であるため、この差分画像の画素変化面積が大きくかつ画像全体に広がることになる。図3（e）の場合、カットとなる可能性が大きい。一方、カメラのパンによる動きの場合変化は小さく、被写体の動きなどは局所的にしか変化しない。ズームでは、広い範囲で変化が起こるが、時間軸上で連続した変化となるので、カットの判定とは見分け

ることが可能である。この画素変化面積の変化を時間とともに示したものが図4である。変化量がしきい値より大きく、なおかつ1フレームにおいて大きな変化が起こる点をカットであると判定して、カットとして検出する。

【0024】次に、シーン検出手段103の処理について説明する。シーン検出手段103は、カット検出手段102によりまとめられたショット群間の類似度を求め、この類似度からシーンにまとめるものである。これによって類似度が所定の条件より低いショット群間をシーンチェンジとし、ショット群をシーンとしてまとめることができる。前記類似度として、例えば、色の分布をもとにしたものがある。これは、フレームを領域分割し、各領域の色とその色の分布比率から、フレーム間の画像間距離を求め、その距離が特定の閾値より大きければ、シーンチェンジとするものである。その他、輝度を用いたものや、動きを求める方法でも適応可能である。

【0025】色の分布を用いる方法について、図5を用いて説明する。まず、ショットの中から1枚代表画像として選び、その画像を色を基にクラスタリングすることで領域分割する。クラスタリングした領域の面積とその色の分布比率をグラフにしたものが図5下の色分布グラフである。この色の種類と分布比率から画像間の距離を求めることで、類似度を計測する。図5（a）と図5

（b）とは、色の種類数が同じで、各色の分布比率も所定の範囲内であるので、類似度が高く、シーンチェンジでないと判断され、図5（b）と図5（c）とは、色の種類数が異なり、色の分布比率も所定の範囲外であるので、類似度が低く、シーンチェンジであると判断される。

【0026】次に、統合手段104の処理について図6を用いて説明する。図6は、連続する映像をシーン単位で示したものである。各シーンは図の様にフレーム群で構成されている。シーン1、シーン2、シーン3、シーン4、シーン5は、シーン検出手段103によって検出されたもので、 $t_1$ 、 $t_2$ 、 $t_3$ 、 $t_4$ 、 $t_5$ はシーンの先頭時刻を示している。また、統合手段104は、相対時刻計算手段106からの時間情報により所定の時間 $T$ ごとに区切りを設け、区切りの時刻におけるシーン番号を求め、その次のシーン番号から新しいシーンとして統合する。つまり、統合手段104によって、新しいシーン（シーン10、シーン11）として統合するのである。この時、シーン1、シーン2、…は、上位シーン10、11の下位シーンとして、記憶部105に階層的に保存される。以下、同様に時間 $T$ ごとにシーンの統合処理を行なうことにより、類似した映像が、極端な時間長でなく適切な時間長でシーン統合がなされ、検索や編集に便宜を図ることができる。

【0027】次に、記憶部105について説明する。記憶部105は、ハードディスクなどの記憶装置である。

この記憶部105に記憶されるデータ構造例は、図8で示され、カット検出手段102より検出されたショットの先頭フレーム番号を記憶する領域、シーン検出手段103により検出された下位シーンの先頭フレーム番号を記憶する領域、統合手段104により検出された上位シーンの先頭フレーム番号を記憶する領域からなる。図8の例では、アドレス0000～に上位シーン領域、0010～に下位シーン領域、0100～にショット領域が確保されている。

【0028】以上の処理方法を図7のフローチャートを用いて説明する。はじめに、記憶領域を確保し、初期設定する(S100)。統合の基準になる先頭フレームの相対時刻を設定する(S101)。次に、画像を処理装置内のメモリ上に取り込む。この時画像のフレーム番号も読み込む(S102)。前フレームとの比較により、カットであるか判定する(S103)。カットと判定されれば、記憶部105のショット領域アドレス上にショット先頭番号を入れる(S104)。さらに、このカットが、シーンチェンジであるか判定し(S105)、シーンチェンジであれば、記憶部105の下位シーン領域アドレス上にショット先頭番号を入れる(S106)。それと同時に、先頭フレームからの順番から入力フレームの相対時刻を求める(S107)。この相対時刻と指定時刻Tを比較し(S108)、越えておれば、そのカットまでを1つのシーンとして統合し、記憶部105の上位シーンアドレス上に先頭フレーム番号を記憶する(S109)。指定時刻Tを越えた時の先頭フレームの相対時刻を0にセットする(S110)。入力が終了(S111)すれば終了し、そうでなければ次の映像を読み込む(S102)。

【0029】(第2の実施の形態)図9に、第2の実施の形態の映像処理装置の構成を示す。第1の実施の形態と相違するのは、構成上、CMを検出するCM検出手段201を設けた点であり、それ以外の構成は、第1の実施の形態と同様であるので説明を省略し、CM検出手段201に関する箇所について説明する。

【0030】CM検出手段201は、多くのCMが60秒以下であって、所定時間である15秒の数倍(1、2、4倍等)時間であること、CMとCMの継目では、シーンチェンジが必ずあること、また複数のCMが続いて繰り返されることが多く、CM中はカットが多いことを用い、カット間の時間差を計算することでCMを検出する。

【0031】図10に、連続する映像をCM検出手段201で得られるCMシーンとそれ以外のショット群を示す。相対時刻計算手段106によって、CMシーン1、2の開始時刻および終了時刻を検出し、図10では、 $t_1$ 、 $t_2$ がCM開始時刻、 $t_1'$ 、 $t_2'$ が終了時刻である。統合手段104は、検出したCM終了時刻から次開始時刻までをショット群1、2、3をシーン30、3

1、32として統合している。

【0032】次に、本実施の形態の処理方法を図11のフローチャートを用いて説明する。はじめに、画像を処理装置内のメモリ上に取り込む(S201)。前フレームとの比較により、カットであるか判定する(S202)。それと同時に、入力されたフレームの順番から先頭フレームからの順番相対時刻を求める(S203)。カットと判定されれば、記憶部105のショット領域アドレス上にショット先頭番号を入れる(S204)。相対時刻が15秒の整数倍であれば、その間はCMであったと判定し、記憶部105のCMシーン領域アドレス上にCMシーンの先頭フレーム番号を入れる(S207)。そして、CMとCMの間を1つのシーンとし、記憶部105の下位シーン領域アドレス上に先頭番号を入れる(S208)。15秒の倍数でなければ、相対時刻が設定時間を越えているかどうか判定する(S206)。例えば、CMは60秒以内とし、この時間を過ぎているかどうか判定する。越えていれば、相対時刻を0にリセットする(S209)。入力が終了(S210)すれば終了し、そうでなければ次の映像を読み込む(S201)。

【0033】ここで、記憶部105に記憶されるデータ構造例は、図12で示され、カット検出手段102により検出されたショットの先頭フレーム番号を記憶する領域、シーン検出手段103により下位シーンの先頭フレーム番号を記憶する領域、統合手段104により上位シーンの先頭フレーム番号を記憶する領域、CM検出手段201によりCMの先頭フレーム番号を記憶する領域からなる。このように領域を区別することで、上位シーンでは、CMが除去された映像が得られ、CMと他の映像とを区別して表示手段に表示でき、検索や編集に便宜を図ることができる。

【0034】なお、本実施の形態では、検出されたCMを区切りとしてその間のショット群をすべて統合処理したが、第1の実施の形態と同様に、シーンチェンジを区切りとして統合するのを組み合わせてもよい。

【0035】(第3の実施の形態)図13に、第3の実施の形態の映像処理装置の構成を示す。符号101～103の構成は、第1の実施の形態と同様であるので説明を省略する。図13において、301は画像入力端101から入力される映像と同期して音声が入力される音声入力端であり、302は音声の特徴となる参照パターンを抽出するパターン抽出手段であり、303は入力された音声と参照パターンとを比較する音声類似度計測手段であり、統合手段104は、音声類似度計測手段303により求めた音声類似度とシーン検出手段103により統合したシーンとによりシーンを統合し、記憶部105のデータ構造例は、図示しないが、音声の参照パターンを記憶する領域がさらに設けられている以外は図8と同様である。

【0036】次に、本実施の形態の音声処理について、図14を用いて説明する。映像と音声とが同期して入力され、まず映像のカットが検出されると、そのショット内での音声データを参照パターンとしてパターン抽出手段302により抽出し、記憶部105に記憶させる。音声類似度計測手段303は、その後入力される音声と記憶されている参照パターンとを比較して、音声類似度を求め、カットの前後において、類似度が小さければ、シーンチェンジと判定し、統合手段104はその判定結果でシーンを統合する。この音声データの類似度計測は、一定期間の音声データ（パターン）を周波数に変換し、周波数軸でパターンマッチングをとることで求めればよい。また、シーンチェンジと判定されれば、参照パターンを記憶部105に新たに記憶させて、以降の比較に用い、シーンチェンジと判定されるたびに新たに記憶される。

【0037】本実施の形態の処理方法を図15のフローチャートを用いて説明する。はじめに、画像および音声処理装置内のメモリ上に取り込む（S301）。前フレームとの比較により、カットであるか判定する（S302）。カットと判定されれば、記憶部105のショット領域アドレス上にショット先頭番号を入れ（S303）、さらに参照パターンをショットに対応の音声データから取り出し、記憶部105に保存する（S304）。記憶部105に保存された参照パターンと以降入力されて検出されたパターンとを比較し、類似度を求め、類似しているかどうか判断し（S305）、シーンチェンジであると判定すると、シーン領域アドレス上にフレーム番号を保存する（S306）。入力が終了（S307）でなければ、画像と音声を読み込み（S301）、以下同様に処理する。

【0038】このようにして、音声の特徴によりシーンチェンジを検出してシーン統合をすることができるため、画像の類似度のみでシーン統合した場合よりもより適切なシーン統合ができ、検索や編集に便宜を図ることができる。

【0039】（第4の実施の形態）図16に、第4の実施の形態の映像処理装置の構成を示す。符号101～103の構成は、第1の実施の形態と同様であるので説明を省略する。図16において、401は図示しない指示手段により記憶部105に記憶されている映像の各階層のショット数を指示する各階層ショット数決定手段、402はショット数決定手段401で指示されたショット数の階層構造となるように記憶部105の内容を階層化する階層構造構成手段、403は入力される音声と画像の特徴量を検出する特徴量検出手段、404は階層構造構成手段402により作成された階層及びショット数で先頭フレームを表示する表示再生手段である。

【0040】本実施の形態では、長い映像全体を素早く把握するために、最初に映像全体を記憶部105の上位

シーンの先頭フレームを表示再生手段404に表示し、さらに詳しく知りたい範囲、つまり、その先頭フレームを指示すれば、その指示されたフレームに対応する下位シーンの階層の映像を指示されたショット数で表示再生手段404に表示するのである。例えば、表示再生手段404に4枚の画像が表示できる場合、映像全体を上位シーンとして4つのショット数に特徴量検出手段403で検出された特徴量を基に分割し、ショット数が4以下になるまで繰り返すのである。また、同様にして、下位シーン領域の映像をショット数が4以下になるまで分割するのである。また、本実施の形態では、ショット数を4としたが、この数に限定されるものでなく、また各階層でのショット数を下層ほど多くする等に変化させてもよい。

【0041】次に、本実施の形態において、ショットとその特徴量を取り込んでから階層構造を再構築する処理を図17のフローチャートを用いて説明する。はじめに、画像および音声処理装置内のメモリ上に取り込む。この時画像のフレーム番号も読み込む（S401）。画像および音声の特徴量を検出する（S402）。前フレームとの比較により、カットであるか判定し、カットと判定されれば、記憶部105のショット領域アドレス上にショット先頭番号を入れ（S403）、ショットごとに記憶部105の特徴量記憶領域に特徴量を保存する（S404）。以下、入力終了（S405）まで繰り返し行なう。これらの処理によりショットと対応の特徴量の記憶が終了する。

【0042】入力終了後、所望の映像を得るための実際の表示操作となる。画面に表示可能なフレーム数より各階層のショット数を各階層ショット数決定手段401により設定する（S406）。ここでは、各階層ともショット数を4に設定した。すると、最初の表示時に、近い特徴量を有する記憶部105のショット同士を設定されたショット数のグループに分割して上位シーンとして表示する。次に、記憶部105に設定数分の下位シーンデータ領域を作成する（S407）。記憶部105からS404で作成した特徴量データを取り出す（S408）。指定された範囲の全てのショットを調べ続け（S409）、取り出した特徴量に基づいて映像を設定数で分割する（S410）。記憶部105の下位シーン領域アドレス上に先頭フレーム番号を保存する（S411）。分割後のすべてショット群において、ショット数が設定数より少なければ終了（412）し、そのショット群内についてS407以降を繰り返す。以下、同様にして、各階層のショット数が4になるまで分割する。このように記憶部105を分割して再構成することで、表示、検索、編集の便宜を図ることができる。

【0043】この記憶部105に記憶されるデータ構造例は、図18で示され、図8と同様、ショットの先頭フレーム番号を記憶する領域、下位シーンの先頭フレーム

番号を記憶する領域、上位シーンの先頭フレーム番号を記憶する領域で構成され、各層において、設定数（図18では4）のショット群が統合されて記憶されている。さらに、各ショットごとの特徴量を記憶する特徴量記憶領域が設けられている。

【0044】（第5の実施の形態）図19に、第5の実施の形態の映像処理装置の構成を示す。本実施の形態は、映像の中から特定の特徴量を持つショットを指定した階層から取り出すものであり、符号101、102は第1の実施の形態と同様であり、符号301、403は第4の実施の形態と同様であるため、説明を省略する。501は対象とする特徴量を設定する特徴量設定手段、502は特徴量設定手段501で設定した特徴量と記憶部105から取り出した特徴量とを比較して類似度を検出する類似度検出手段である。

【0045】次に、本実施の形態において、ショットとその特徴量を取り込んでから所望の特徴量をもつショットを取り出す処理を図20のフローチャートを用いて説明する。はじめに、画像および音声処理装置内のメモリ上に取り込む。この時画像のフレーム番号も読み込む（S501）。画像および音声の特徴量を検出する（S502）。前フレームとの比較により、カットであるか判定し、カットと判定されれば、記憶部105のショット領域アドレス上にショット先頭番号を入れ（S503）、ショットごとに確保された特徴量記憶領域に特徴量を保存する（S504）。以下、入力終了まで（S505）繰り返し行なう。これらの処理によりショットと対応の特徴量の記憶が終了し、検索の準備が整う。

【0046】入力終了後、所望の映像（ショット）を得るために、検索する特徴の条件を設定する（S506）。すると、記憶部105から特徴量データを取り出し（S507）、取り出した特徴量データが検索の条件を満たすなら（S508）、表示のために、下位または上位のシーン領域アドレス上に先頭フレーム番号を保存する（S509）。以下、全てのショットについて調べる（S510）。別の特徴でも条件を設定するなら（S511）、S506に戻り繰り返す。

【0047】この記憶部105に記憶されるデータ構造例は、図21で示され、図8と同様、ショットの先頭フレーム番号を記憶する領域、下位シーンの先頭フレーム番号を記憶する領域、上位シーンの先頭フレーム番号を記憶する領域で構成されている。さらに、各ショットごとの特徴量を記憶する特徴量記憶領域が設けられている。この特徴量記憶領域は、図21に示す如く、特徴ごとに設けられていてもよい。

【0048】図22に、特徴量検出手段403の第1の具体例を示す。図22では、画像入力端101から入力された画像は、フロー検出処理601で複数枚のフレーム間でオブティカルフローを計測する。カメラ方向計測処理602では、フロー検出処理601で得られたフロ

ー結果を基にカメラの動きを判定する。

【0049】オブティカルフローとは、画面内の物体の移動に伴う輝度分布の移動を示す速度ベクトルである。このオブティカルフローの検出方法としては、大きく分けて二種類ある。一つは、画像中の特徴点を見つけ、画像フレーム間で対応点を探し速度ベクトルを決定するマッチング法であり、もう一つは、動画画像中の運動物体の輝度の時間変化に一定の関係があることを利用する勾配法である。

【0050】上記で計測したオブティカルフローの分布とカメラの動きの関係は、図23で示される。カメラがパン・チルトすると左右あるいは上下の1方向のベクトルが得られ（図23（a）、図23（b））、ズームあるいはワイドの時は画像の中心から放射状のベクトルが得られる（図23（c））。パンとチルトの混合した場合は、平行のベクトルが得られる（図23（d））。パンとチルトおよびズーム・ワイドが混合した場合は、画像の中心ではないある点を中心に放射状のベクトルとなる（図23（e））。

【0051】また、カメラの動きと画像上のベクトルの関係は下式で表現される。

【0052】

【数1】

$$\begin{pmatrix} u \\ v \end{pmatrix} = a \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} p_x \\ p_y \end{pmatrix}$$

【0053】ここで、 $(u, v)$ は動きベクトル、 $x, y$ は画面座標値、 $a$ はズーム成分、 $p_x$ はパン成分、 $p_y$ はチルト成分の係数を表す。上式と図23との関係は、図23（a）は $a=0, p_x \neq 0, p_y=0$ を満たす。図23（b）は $a=0, p_x=0, p_y \neq 0$ を満たす。図23（c）は $a \neq 0, p_x=0, p_y=0$ を満たす。図23（d）は $a=0, p_x \neq 0, p_y \neq 0$ を満たす。図23（e）は $a \neq 0, p_x \neq 0, p_y \neq 0$ を満たす。

【0054】カメラの動きの判定は、まずベクトルが平行であるかを判定する。平行であれば $a=0$ が決定し、 $p_x, p_y$ もベクトルから求まる。平行でない場合、カメラの中心（E OF）を求め、その中心からの距離とベクトルの大きさの関係よりカメラのズーム成分を求める。これによって $a$ が得られる。その後、上式に $a$ を代入することで、 $p_x, p_y$ が求まる。この3成分 $a, p_x, p_y$ が決まれば、カメラの動きが決定され、これらを、映像の特徴量として用いる。このカメラの動きでショットを表示、あるいは検索することが可能となる。

【0055】図24に、特徴量検出手段403の第2の具体例を示す。図24では、音声入力端301から入力された音声は、音量検出処理部701において、音の大きさを測定し、音量記憶手段702に記憶させ、後から入力される音声との変化を順次検出する。この方法によ



って、音量変化が大きい、あるいは音量自体が大きい場合を特徴量とする。これにより、音量変化の大きいショットあるいは音量が大きいショットを表示、あるいは検索することが可能となる。

【0056】図25に、特徴量検出手段403の第3の具体例を示す。図25では、ある特定の環境下における特定の形状などの典型的な画像例を物体モデル（画像モデル）801として持ち、そのモデルを基に画像入力端101から入力される画像との類似度を類似度計測手段802で計測し、求めた類似度を特徴量として用いる。物体モデルとしては、例えば人の顔、車、風景など様々なものを用いることが可能である。これにより、物体モデル801に近いショットを表示、あるいは検索することが可能となる。

【0057】図26に、特徴量検出手段403の第4の具体例を示す。図26では、ある特定の環境下におけるノイズなどの典型例を環境モデル（音声モデル）901として持ち、そのモデルを基に音声入力端301から入力される音声との類似度を類似度計測手段902で計測し、求めた類似度を特徴量として用いる。環境モデルとしては、例えば車内の雑音、列車音など様々なものを用いることができる。これにより、環境モデル901に近いショットを表示、あるいは検索することが可能になる。

【0058】以上、種々の実施の形態について説明してきたが、これらの実施の形態を適宜組み合わせ用いてもよく、さらには、上述した処理を実行するプログラムを、フロッピーディスクやCD-ROM等のコンピュータ読み取り可能な記録媒体にあらかじめ記録したり、あるいは通信回線でコンピュータ読み取り可能な記録媒体に供給し、適宜コンピュータにインストールして用いてもよい。

#### 【0059】

【発明の効果】請求項1記載の発明によれば、長い映像を意味のある短いシーンにまとめることができるので、映像内容の理解が簡単になる。

【0060】請求項2記載の発明によれば、CMと映像内容とを区別して分かりやすくまとめることができるので、映像内容の理解が簡単になる。

【0061】請求項3記載の発明によれば、映像のみでシーン統合した場合よりも意味的に適切なシーン統合ができるので、映像内容の理解が簡単になる。

【0062】請求項4記載の発明によれば、上記効果に加え、記憶手段のショットを分割して再構成することにより、表示に適切な形にまとめることができ、映像中から重要な部分だけを取り出すことができ、検索、編集の便宜を図ることができる。

【0063】請求項5記載の発明によれば、上記効果に加え、記憶手段から検索条件として設定した特徴に近い所望のショットを記憶手段から検索して取り出すことが

できる。

【0064】請求項6記載の発明によれば、所定の映像中のカメラの動き情報に近いショットを表示、あるいは検索することが可能になる。

【0065】請求項7記載の発明によれば、音量変化あるいは音量の大きなショットを表示、あるいは検索することが可能になる。

【0066】請求項8記載の発明によれば、特定の画像モデルに近いショットを表示、あるいは検索することが可能になる。

【0067】請求項9記載の発明によれば、特定の音声モデルに近いショットを表示、あるいは検索することが可能になる。

#### 【図面の簡単な説明】

【図1】第1の実施の形態の映像処理装置の構成を示す図である。

【図2】本発明における映像の構成を表す概念図である。

【図3】第1の実施の形態のカット検出として画素変化面積を用いる場合の説明図である。

【図4】第1の実施の形態の画素変化面積の時間経過を示す図である。

【図5】第1の実施の形態のシーン検出として色分布を用いる場合の説明図である。

【図6】第1の実施の形態のシーン統合の説明図である。

【図7】第1の実施の形態のフローチャートである。

【図8】第1の実施の形態の記憶部のデータ構造の構成例である。

【図9】第2の実施の形態の映像処理装置の構成を示す図である。

【図10】第2の実施の形態のシーン統合の説明図である。

【図11】第2の実施の形態のフローチャートである。

【図12】第2の実施の形態の記憶部のデータ構造の構成例である。

【図13】第3の実施の形態の映像処理装置の構成を示す図である。

【図14】第3の実施の形態の音声データによりシーンチェンジを検出する説明図である。

【図15】第3の実施の形態のフローチャートである。

【図16】第4の実施の形態の映像処理装置の構成を示す図である。

【図17】第4の実施の形態のフローチャートである。

【図18】第4の実施の形態の記憶部のデータ構造の構成例である。

【図19】第5の実施の形態の映像処理装置の構成を示す図である。

【図20】第5の実施の形態のフローチャートである。

【図21】第5の実施の形態の記憶部のデータ構造の構

成例である。

【図 2 2】特徴量検出手段の第 1 の構成の具体例である。

【図 2 3】特徴量検出手段の第 1 の具体例に関する説明図である。

【図 2 4】特徴量検出手段の第 2 の構成の具体例である。

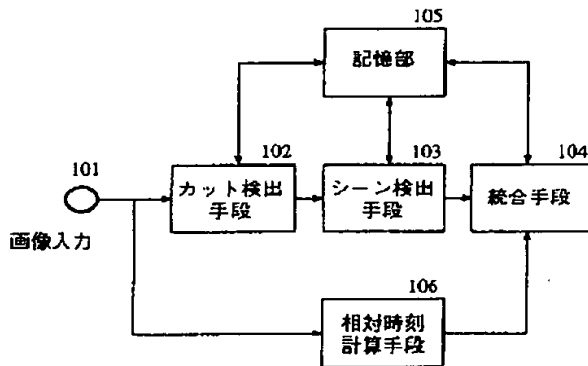
【図 2 5】特徴量検出手段の第 3 の構成の具体例である。

【図 2 6】特徴量検出手段の第 4 の構成の具体例である。

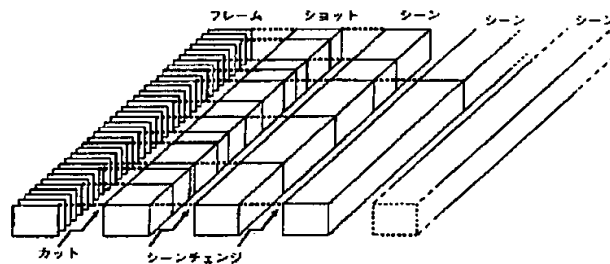
【符号の説明】

- 1 0 2 カット検出手段
- 1 0 3 シーン検出手段
- 1 0 4 統合手段
- 1 0 5 記憶部
- 1 0 6 相対時刻計算手段

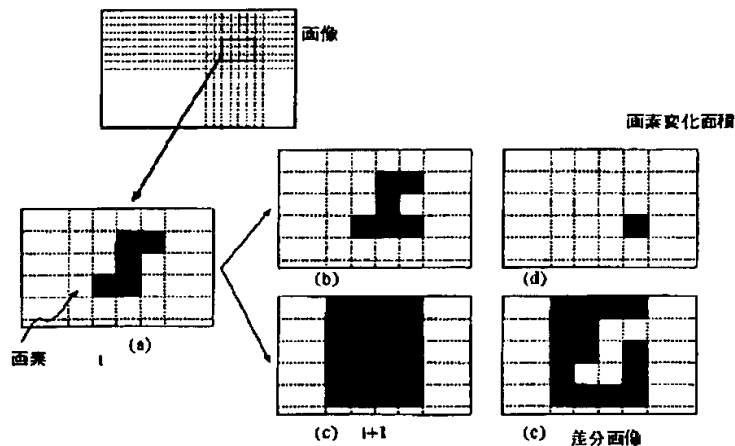
【図 1】



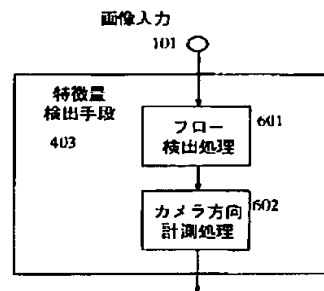
【図 2】



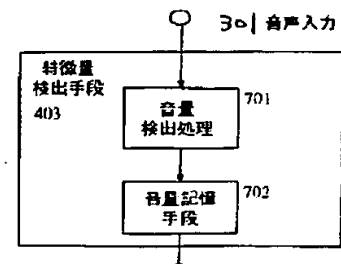
【図 3】



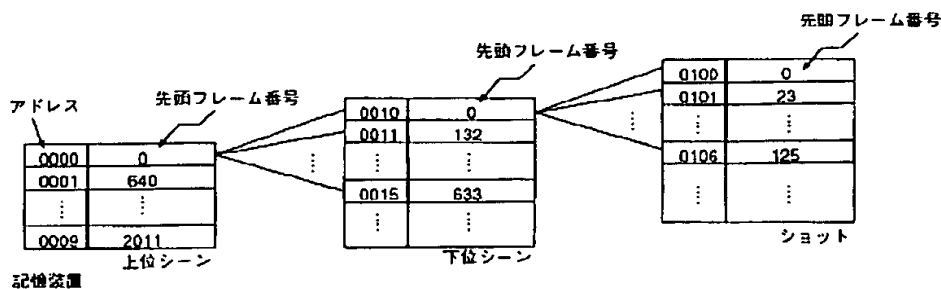
【図 2 2】



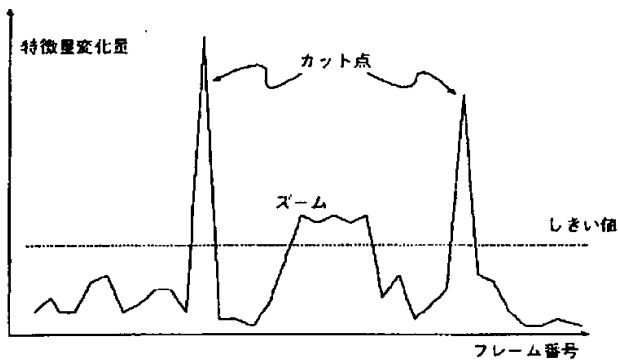
【図 2 4】



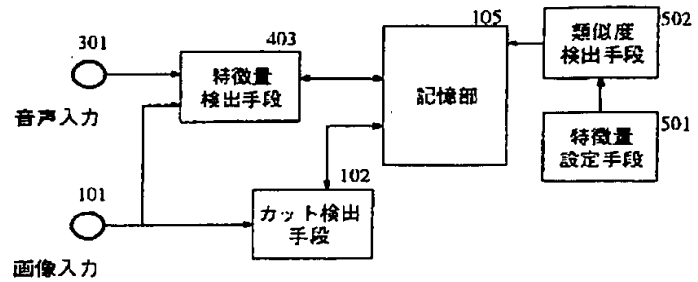
【図 8】



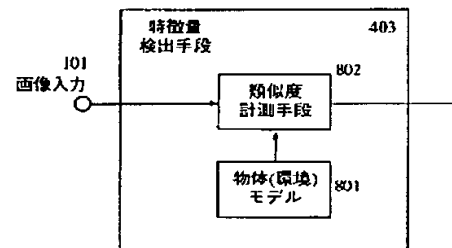
【図4】



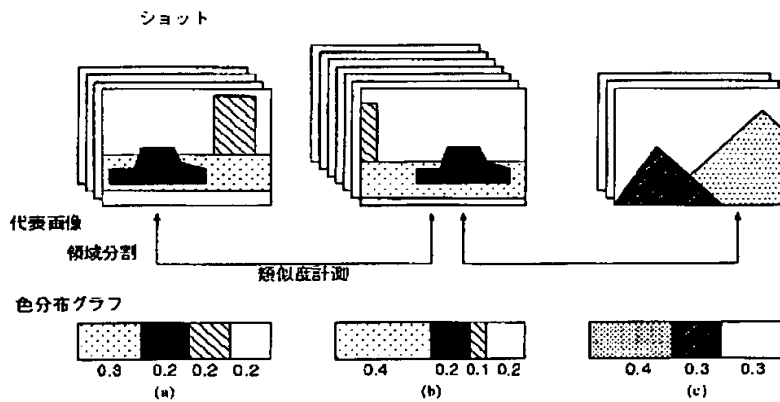
【図19】



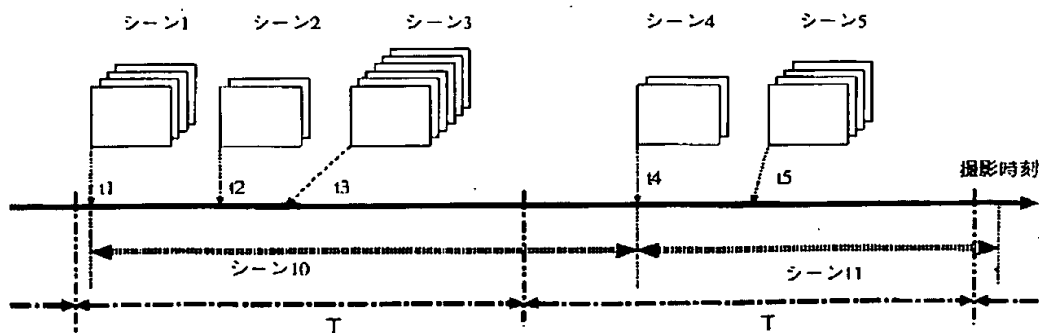
【図25】



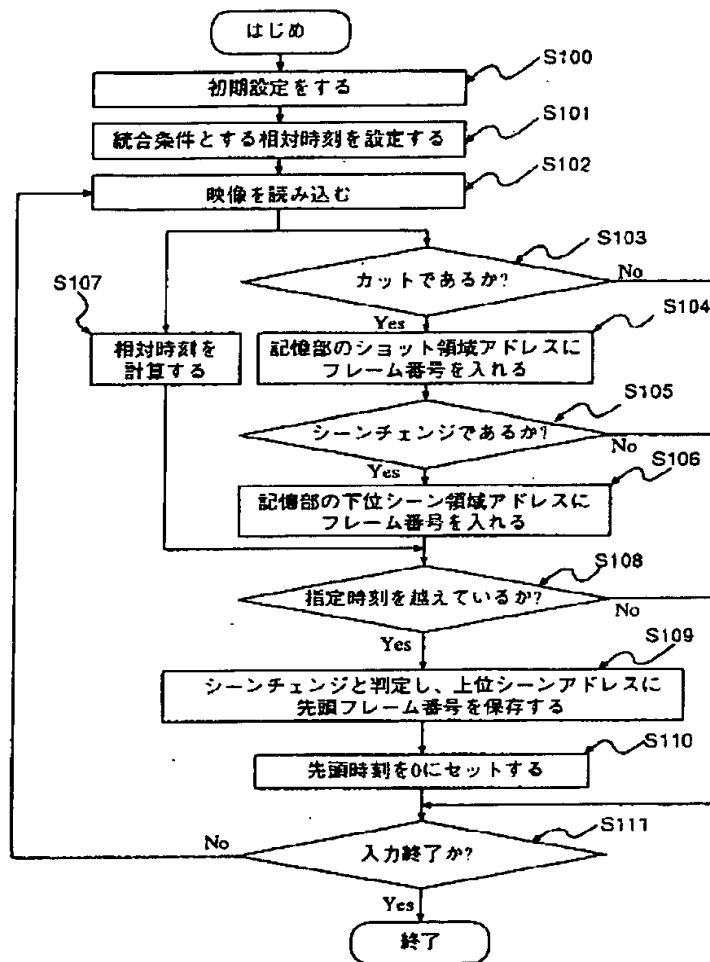
【図5】



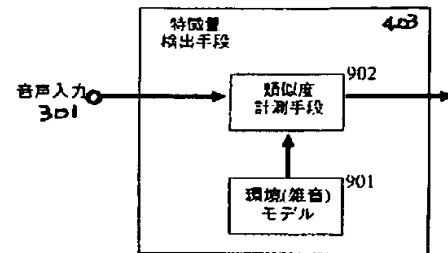
【図6】



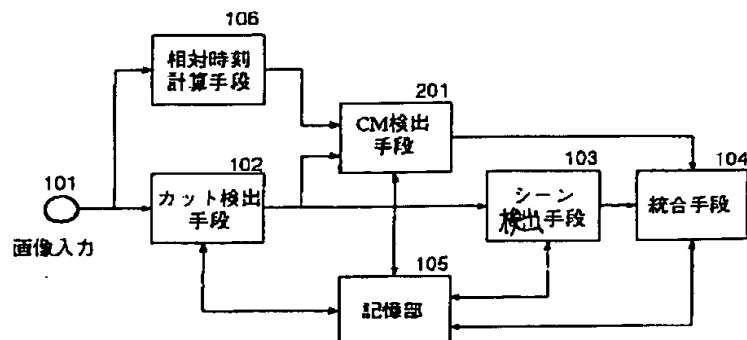
【図7】



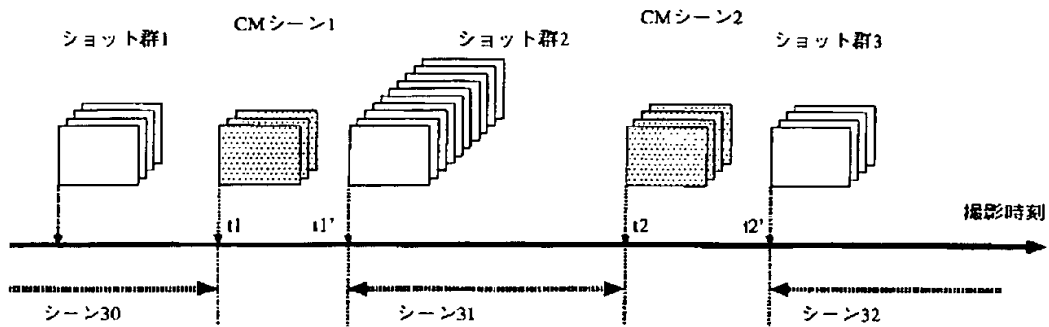
【図26】



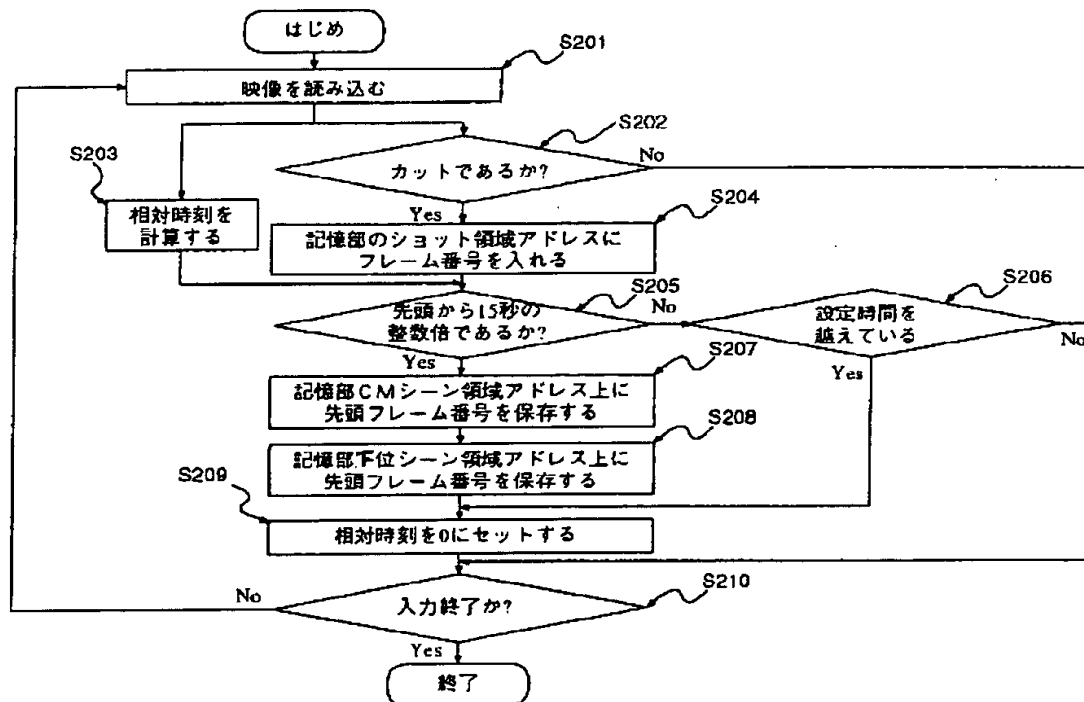
【図9】



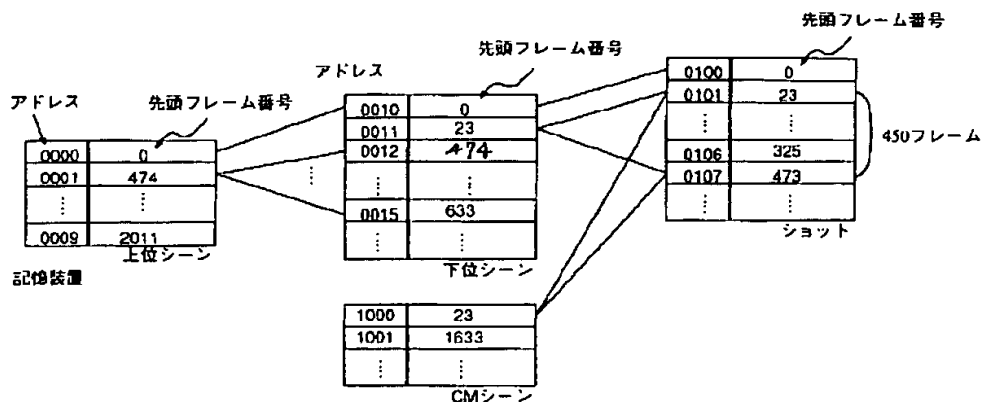
【図10】



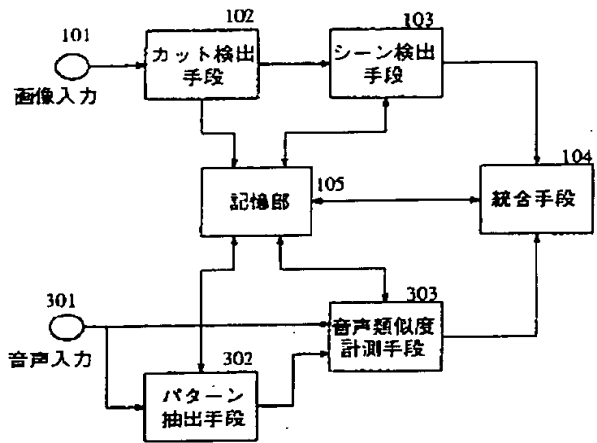
【図11】



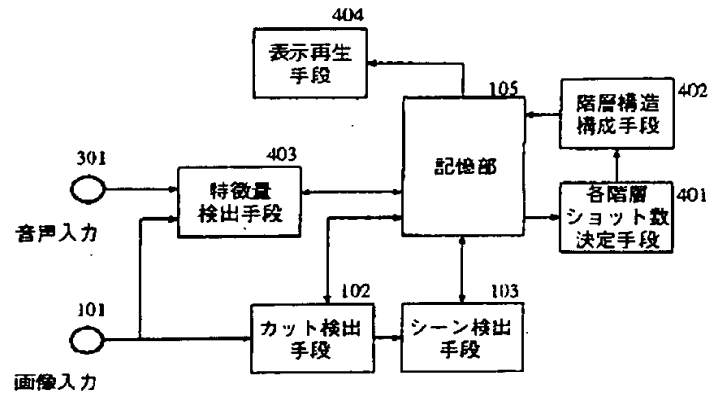
【図12】



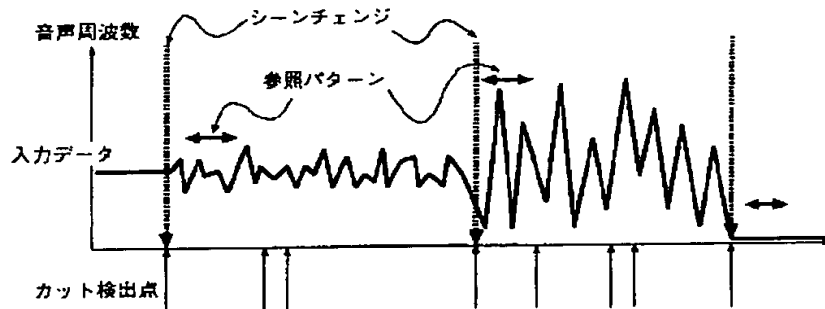
【図13】



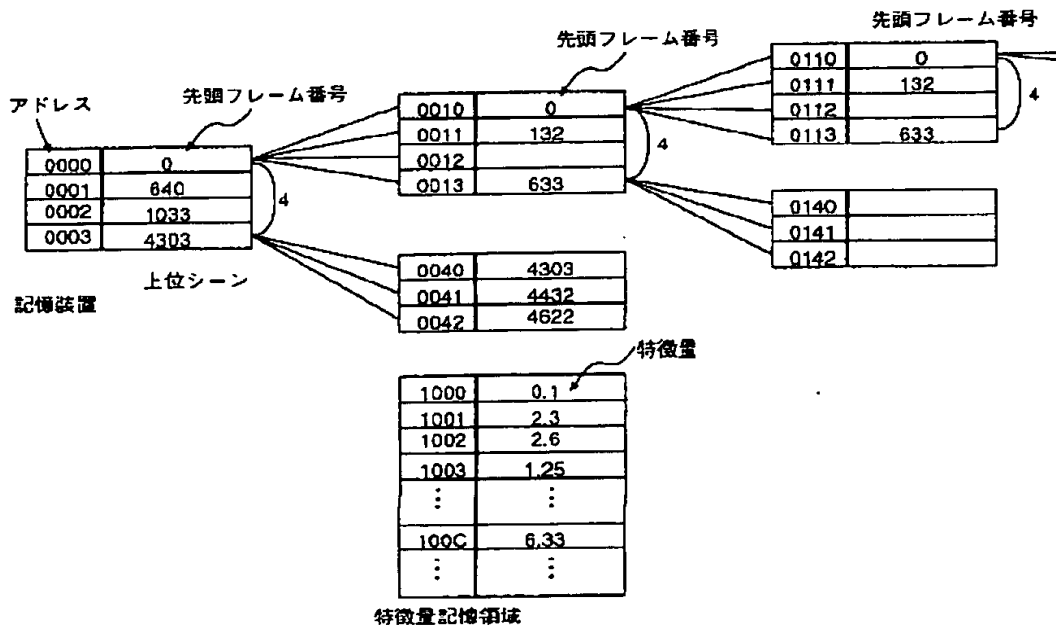
【図16】



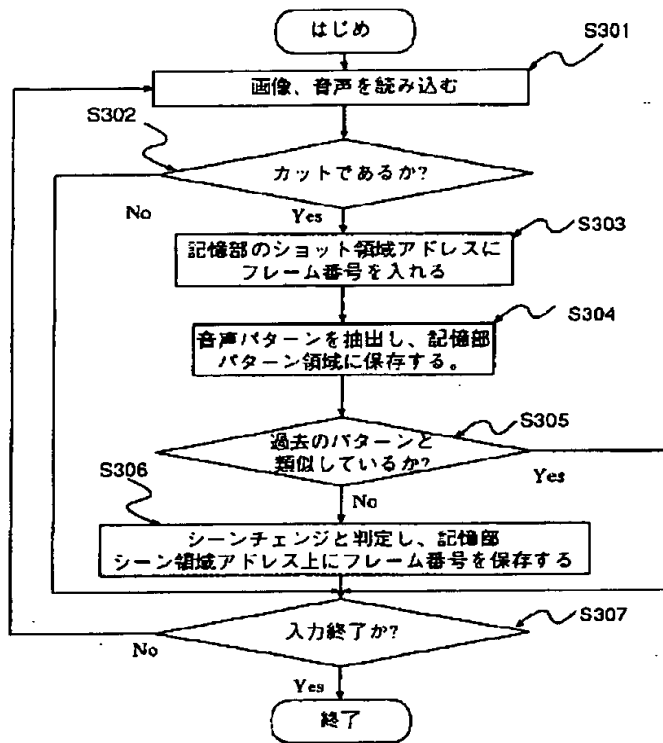
【図14】



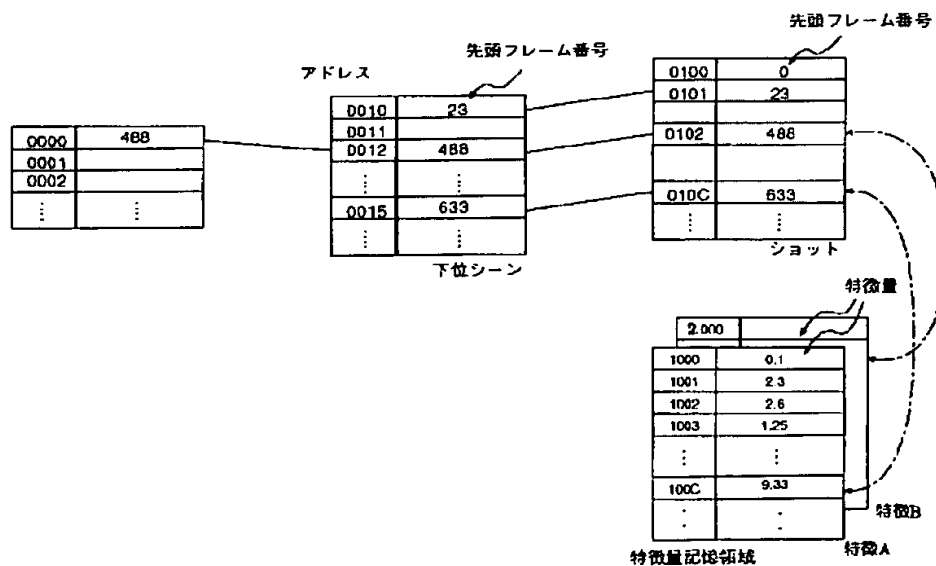
【図18】



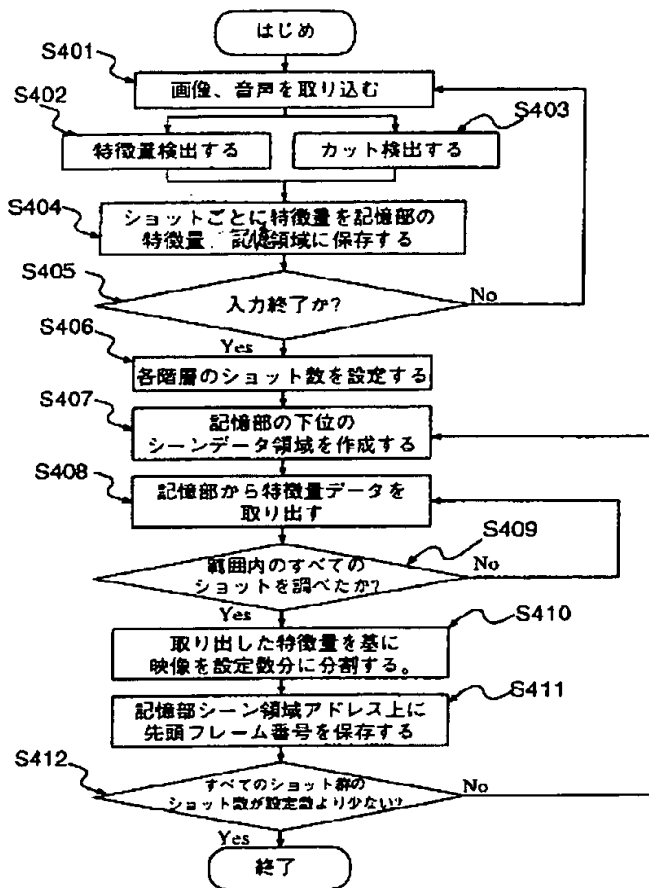
【図15】



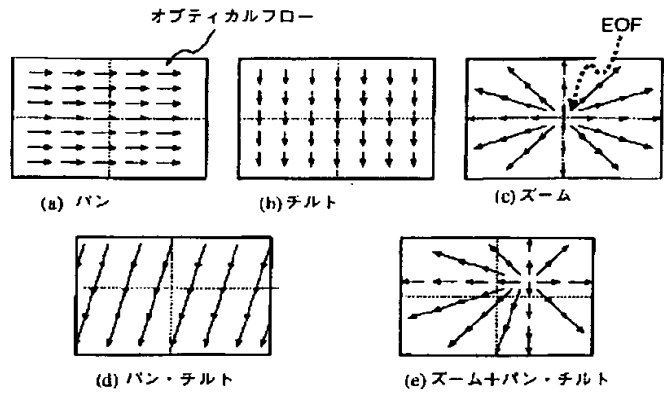
【図21】



【図17】

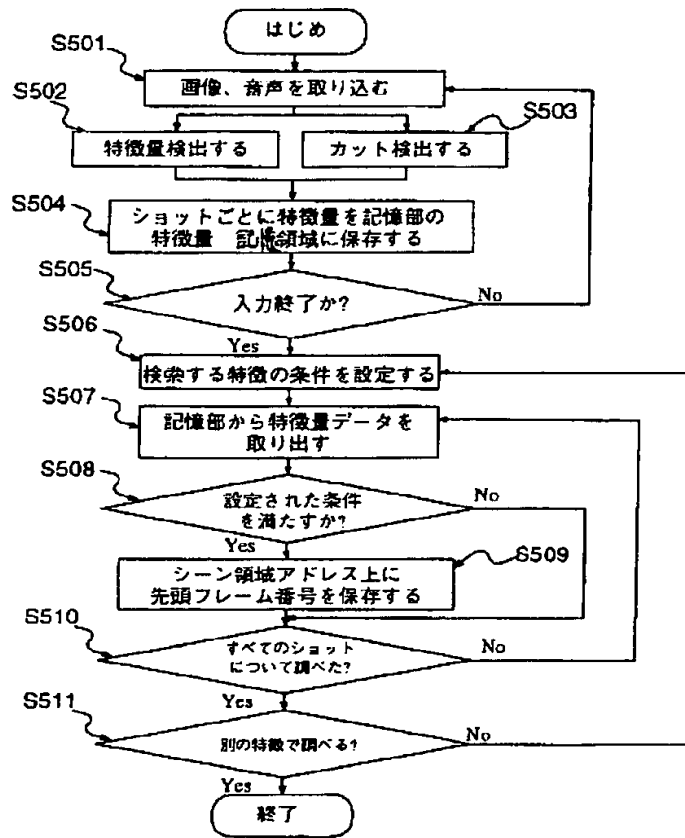


【図23】





【図20】



フロントページの続き

(72)発明者 中村 三津明  
大阪府大阪市阿倍野区長池町22番22号 シ  
ヤープ株式会社内

Fターム(参考) 5C023 AA14 AA34 AA37 AA38 BA02  
CA01 CA04 DA04 EA13  
5C052 AA01 AA17 AC08 CC20 DD06  
5C053 FA05 FA14 FA21 FA23 GB09  
GB19 HA29 HA30 HA40 JA12  
JA24 JA30 KA05